DOCUMENT RESUME

ED 158 834 PS 009 537

AUTHOR Gallimore, Ronald; And Others
TITLE Pretest and Posttest Results of the First KEEP Program Year. Technical Report #5.
INSTITUTION Kamehameha Schools, Honolulu, Hawaii. Kamehameha Early Education Project.
SPONS AGENCY California Univ., Los Angeles. Mental Retardation Research Center.; Hawaii State Dept. of Education, Honolulu.
PUB DATE [73]
NOTE 18p.; For related documents, see PS 009 533-536 and PS 009 538-573

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Demonstration Projects; Hawaiians; *Intelligence Quotient; Intelligence Tests; *Kindergarten; *Program Evaluation; *Socioeconomic Background; Standardized Tests
IDENTIFIERS Hawaii; *Kamehameha Early Education Program

ABSTRACT

This report presents a pre- and posttest evaluation of the first program year of the Kamehameha Early Education Program (KEEP). Each of the 28 children in the KEEP kindergarten class received three tests: the Wechsler Preschool and Primary Scale of Intelligence (WPPSI), the Metropolitan Readiness Test (MRT), and the Standard English Repetition Test (SERT). The testing scores from the fall and spring were compared on the basis of their correlations with each other and with other variables, such as socioeconomic factors. Improvement on the MRT was only moderately correlated with changes in I.Q. scores. The children with the greatest changes in I.Q. scores tended to have had less school experience and came from lower income families. The high and low MRT change groups showed significant differences on only three of 100 variables, and the three were not in any particular group of variables, as were significant variables of the high and low I.Q. change groups. SERT scores were highly correlated with all three I.Q. test measures for both pre- and posttestings, and were consistently related to socioeconomic factors. While SERT scores did significantly increase from fall to spring, the change in SERT scores did not correlate with initial scores. (Author/JMB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED158834

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Technical Reports

of

The Kamehameha Early Education Program

a research and development program established and funded by

The Kamehameha Schools/Bernice P. Bishop Estate

Ronald Gallimore, Roland G. Tharp & Gisela E. Speidel,
General Editors

Pretest and Posttest Results of the First KEEP Program Year.

Ellen Antill
Production Editor

Technical Report #5

The cooperation of the State of Hawaii Department of Education is gratefully acknowledged, as is the support and resources made available by the Sociobehavioral Research Group, MRRC, University of California, Los Angeles.

PS 009537

The opinions expressed herein do not necessarily reflect the position, policy or have the endorsement of The Kamehameha Schools/Bernice P. Bishop Estate, or of the editors.

Published by The Kamehameha Early Education Project, 1850 Makuakane Street, Honolulu, HI 96817

All rights reserved. No part of this report may be reproduced in any form or by any means without the prior written permission of The Kamehameha Schools/Bernice P. Bishop Estate.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Kim C. M. Sloat

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

The Kamehameha Early Education Program

The Kamehameha Early Education Program (KEEP) is a research and development program of The Kamehameha Schools/Bernice P. Bishop Estate. The mission of KEEP is the development, demonstration, and dissemination of methods for improving the education of Hawaiian and Part-Hawaiian children. These activities are conducted at the Ka Na'i Pono Research and Demonstration School, and in public classrooms in cooperation with the State Department of Education. KEEP projects and activities involve many aspects of the educational process, including teacher training, curriculum development, and child motivation, language, and cognition. More detailed descriptions of KEEP's history and operations are presented in Technical Reports #1-4.

Abstract

This report presents a pre- and posttest evaluation of the first KEEP program year. The first kindergarten class received three tests: the Wechsler Preschool and Primary Scale of Intelligence (WPPSI), the Metropolitan Readiness Test (MRT), and the Standard English Repetition Test (SERT). The scores were compared on the basis of their correlations with each other and with other variables, such as socioeconomic factors. Interestingly enough, improvement on the MRT is only moderately correlated with changes in I.Q. scores.

The children with the greatest changes in I.Q. scores tended to have had less school experience and came from lower income families. The high and low MRT change groups showed significant differences on only three of 100 variables, and the three were not in any particular group of variables, as were significant variables of the high and low I.Q. change groups.

SERT scores were highly correlated with all three I.Q. test measures for both pre- and posttestings, and were consistently related to socioeconomic factors. While SERT scores did significantly increase from Fall to Spring, the change in SERT scores did not correlate with initial scores.

These relationships between tests and the correlations of test scores with other variables must be verified and elaborated on with future classes.

Technical Report #5

Pretest and Posttest Results of the First KEEP Program Year[1]

Ronald Gallimore Roland G. Tharp Gisela E. Speidel

One can consider many dimensions in program evaluation. At present educational research tends to be more concerned with process analysis and less with outcome evaluation. Determining that a program did or did not succeed at the end of a school year is often of minimal value since such efforts frequently neglect to report what occurred in the classroom.

As important as process analysis is, however, it is also important to place the consequences of any educational effort in broader perspective. Despite their limitations, it remains true that standardized tests provide one useful perspective. Data presented here are unusually restricted since we do not have appropriate control groups on major testing devices. Such data could not be collected until we had begun our work with public schools.

Pretest and Posttest Data

Each of the 28 children at KEEP received the WPPSI in Fall, 1972. Later in the Fall of the same year, the Metropolitan Readiness Test (MRT) was also administered. Graduate students in psychology administered both the tests.

The means for the Fall and Spring administrations of the WPPSI and the MRT are presented in Table 1. The considerable variability within the group

---

[1]Preparation of this report was partially supported by the Socio-behavioral Research Group, Mental Retardation Research Center, UCLA. Computing assistance was obtained from Health Sciences Computing Facility, UCLA, supported by NIH Special Research Resources Grant RR-3.

of 28 kindergarteners reflects the heterogeneity of the class. For example, the standard deviations for the Fall Full Scale I.Q. and the MRT raw scores were 20.2 and 16.0, respectively. In Spring, the variation in I.Q. scores is somewhat less than it was in Fall, but the standard deviation for the Spring administration of the MRT is slightly lower (See Table 1).

Table 1

1972-73 KEEP Kindergarten Class Intelligence and Achievement Test Scores

September and May

	Fall (Sept. 1972)		Spring (May 1973)		
WPPSI*	$\bar{X}$ (N=28)	S.D.	$\bar{X}$ (N=27)*	S.D.	Mean Change (S.D.)
Full Scale	92.1	20.2	101.6	13.8	10.63 (13.07)
Verbal	87.5	19.4	92.7	14.7	6.48 (12.49)
Performance	98.3	19.6	110.9	13.6	13.78 (14.33)
Metropolitan Readiness Test					
	$\bar{X}$ (N=28	S.D.	$\bar{X}$ (N=28)	S.D.	
Total Raw Score	32.3	16.0	59.6	11.9	26.69 (11.04)
Percentile	19.0	21.7	60.3	22.3	40.73 (18.26)

*N=27 (one child was not retested on the WPPSI because she left school early)

For both the three I.Q. scores and the MRT, statistically significant increases were obtained from the pre- to the posttesting (the mean change scores are also presented in Table 1). However, the total group I.Q. and achievement test scores conceal an important feature of the results: the change scores actually reflect two essentially different populations. Of the 27 children retested on the WPPSI in May, 14 showed an increase in Full Scale

I.Q. of 10 or more scale points; the mean increase in I.Q. for these 14 children was 20.7. The remaining 13 children showed a mean increase of five I.Q. points. Eight of this group increased their scores less than ten points, and five decreased an average of seven points.

Table 2

I.Q. Change Scores - High and Low Change Groups

	SEPTEMBER 1972		MAY 1973	
	Lo (9 or less pts.) Change (N=27)	Hi (10 or more pts.) Change	Lo Change	Hi Change (N=27)
Full Scale	104.23	73.27	104.46	94.55
Verbal	98.46	71.27	96.08	85.18
Performance	109.31	80.55	112.69	105.91

Table 2 presents the mean pre- and posttest I.Q. scores for the Hi and Lo change groups. In September, the differences are large and dramatic: the Lo change group at the time of the pretest is in the average range of I.Q. scores, while the Hi change group is well below average. At the time of the posttest, however, the gap has narrowed to the point that there are no statistically reliable differences between the Hi and Lo change groups.

It is not possible to rule out regression effects in this case. However, since the KEEP children were selected randomly, the threat to validity posed by regression effects is correspondingly reduced. Consideration of these results should nonetheless be undertaken with appropriate caution.

Dramatic increases also were obtained for the MRT. The actual pre- and posttest scores of 12 children are presented in Table 3. We present the actual scores to illustrate that the changes obtained are not attributable to a few extreme scores; the 12 children showed a mean increase on the MRT posttest of 45 or more points.

Changes on the MRT were only modestly correlated with changes on the WPPSI. The correlations between the MRT and the Full Scale, Verbal, and Performance are: .32, .32, and .18 (df=27), respectively. These insignificant correlations are due to the fact that there was much greater variability in I.Q. change scores than in MRT scores; that is, all children showed improvement on the MRT, but only about half of the group improved on the I.Q. Thus if we were to examine only children who showed improvement on both, there would be a very large correlation between change scores on the two tests.

Again it is prudent to invoke a word of caution about the problem of regression effects. On any psychological or educational measure, it must be expected that changes will tend to move toward the population mean score. In this case, about half of the children tested very low in the Fall. Since there is much more "room" to improve than to score lower, the probability that a "gain" will occur is correspondingly increased. The question always is whether any amount of the observed change in scores can be attributed to some factor of interest, in this case exposure to the KEEP program. We are not prepared to draw that conclusion, but it is appropriate to conclude that the average performance level of the children in Spring is within statistically normal limits, and thus programmatically useful. Whether the change from Fall can be attributed to our efforts is problematical. Continued use of random sampling, replication of these gains with subsequent classes, specific hypothesis testing, continuous monitoring of educational process variables, among other things can all serve to reduce the probability that regression effects are the sole cause of the obtained changes.

Table 3

Pre-, Post-, and Change MRT Scores for 12 Children

who increased scores 45 points or more

September (Percentile)	May (Percentile)	Change (Percentile)
Entering Metro		
27	79	52
08	75	67
13	86	73
03	67	64
08	65	57
08	65	57
07	53	46
05	53	48
13	63	50
05	59	54
11	65	54
25	79	54

Characteristics of Hi and Lo Change Groups

Those children who showed the greatest increase in Full Scale WPPSI I.Q. scores came from families in which the combined occupational status score of the parents and of the mother alone was relatively lower. Also, children who showed most I.Q. change tended not to have attended preschool (See Table Q).

Table Q

Significant Correlates of I.Q. Full Scale Change Scores
Class 1, 1972-73

Parent's Occupation	-.47 (n = 27)
Mother's Occupation	-.43 (n = 27)
Preschool	-.40 (n = 25)

In Table 4, the mean I.Q. Change scores for children attending and not attending preschool are presented. Although those who did not attend preschool showed a significantly greater increase on Full Scale I.Q., the principal contribution to the difference between the groups appears to be the differential score on Performance I.Q. Thus, children who did not attend preschool showed a significant increase in Performance I.Q., but not Verbal I.Q.

Table 4

Mean I.Q. Change Scores for

Children Attending and Not Attending Preschool

		Preschool		Significance of Mean Difference	
		Yes	No	df	t
Full Scale I.Q.	Change $\bar{X}$	5.08	14.75	23	2.09*
	SD	(10.72)	(12.38)		
Verbal I.Q.	Change $\bar{X}$	2.39	9.08	23	1.43
	SD	(11.22)	(12.21)		
Performance I.Q.	Change $\bar{X}$	7.15	19.50	23	2.40**
	SD	(13.41)	(12.24)		

* $p < .05$ (two-tailed)

** $p < .025$ (two-tailed)

In summary, the children with the greatest changes in I.Q. scores tended

to have had less school experience and to come from lower income families.

Of the 100 variables examined, only three significant correlates of MRT change scores were obtained. The three were: the degree of child care assumed by older male siblings, the average number of seconds the child attended to another child during an initial experience in peer tutoring, and the average attentiveness to a peer tutor over three observation sessions. Marginal, but not statistically significant, relationships were obtained for age, pretest MRT score, and father's responsibility for child care. Children who increased in MRT scores the most tended to be somewhat younger, have lower scores on the MRT pretest, and come from families in which the father assumes greater responsibility for child care.

Unlike I.Q. change scores, the MRT change scores do not show any consistent association with particular classes of variables, such as indices of socioeconomic status. Indeed, it is difficult to understand why a correlation should emerge between MRT change scores and the attentiveness of children to a peer tutor: the peer tutor data were obtained late in the school year, represent relatively brief periods (a total of three ten minute sessions), and were collected in substantially the same manner as a daily on-task observation code. If a relationship had been obtained between the daily on-task observation measure, it might have been appropriate to interpret the correlation between MRT change scores and attentivenss to a peer tutor. As it is, it is more prudent to consider the correlation a chance finding in a 100 variable matrix.

The Standard English Repetition Test (SERT)

The SERT was devised as a measure of Standard English (SE) performance. A more detailed presentation of validity and reliability has been reported

elsewhere (Day, Boggs, Speidel, Tharp, and Gallimore, 1975). The SERT uses the device of sentence repetition to assess the child's SE performance level. Correlations between SERT performance and I.Q. are significant at the beginning of the school year; the correlations are .51, .53, and .44 (N=28) for the Full Scale, Verbal, and Performance portions of the WPPSI, respectively.

At the end of the school year, posttest scores on the SERT and the WPPSI correlate .70, .76, and .43 (N=27) for Full Scale, Verbal and Performance. Both the SERT and the WPPSI were administered in September and May. The increased association of SERT and I.Q. scores in May suggests that SE performance has a greater effect on child performance after school entry. The level of association at entry into kindergarten is also high, however, confirming the expectation that SE performance is a factor in school achievement from the beginning.

Not surprisingly, SERT scores are consistently related to socioeconomic (SES) factors, including father's employment status, occupational level of mother and father, and father's education. Significant correlations between SERT and SES variables were obtained both at the beginning and ending of the school year. However, two variables--mother's employment status and education--were correlated only with the May posttest SERT scores.

In September, the average SERT score was 9.18 (SD=5.72). This value represents the mean number of correct repetitions out of a possible 29. In May, the mean score was 13.61 (SD=6.62), a statistically significant difference (t=7.85, df=27, $p < .001$ for correlated samples). The gain in SE performance was 48.6 percent, and this in a classroom in which no formal instruction in SE or punishment for speaking HCE occurred.

Unlike I.Q. and achievement test data, changes in SERT scores over the school year do not correlate with initial scores (r=.04, df=26). That is,

the level of performance displayed by a child upon entry into kidnergarten does not relate to increased SE performance as measured in May. Gains in SE performance in May cannot be attributed to initially higher levels of performance in September. Whatever produced an increase in SERT scores occurred after the child began kindergarten.

Table 5 presents the significant correlates of SERT change scores, as well as SERT pre- and posttest scores. Unlike the pre- and posttest scores, change scores correlate with only two socioeconomic variables--both parents working and mother's education. Besides pre- and posttest I.Q. scores, the only other significant correlate of change in SERT performance is the degree of the mother's responsibility for child care, and, in this case, children who improve more on the SERT come from homes in which the mother is relatively less responsible. Thus, children who improve on the SERT have higher I.Q. scores, better educated mothers, both parents working, and are cared for by persons other than the mother. The latter is doubtless a function of the greater likelihood of the mother working.

Table 5 indicates the lack of correlation between SERT, I.Q., and MRT Change Scores. This is true despite the fact that SERT pre- and posttest and change scores correlate with pre- and posttest I.Q. scores; indeed, on the posttest, the SERT and WPPSI scores were even more closely related than on the pretest, as reported above. Thus, the children who show change on the SERT and the children who show change on the WPPSI <u>are not</u> the same children. Although the two tests are correlated, relative shifts in score levels are not. The reason for this apparent contradiction can be found in the relationship between pretest and posttest scores and change scores. For the WPPSI, changes in I.Q. are <u>negatively related</u> to incoming scores; for the SERT, there is no relationship between changes and pretest scores. On the posttest,

however, the picture is reversed: in May, there is no relationship between I.Q. change scores and posttest scores. On the SERT, change scores are correlated with posttest scores. In other words, some students began the school year with very low I.Q. scores, and improved to a point indistinguishable from those who had begun with average I.Q. scores. In terms of SE performance, some students ended the year with significantly higher SERT scores although they had been indistinguishable from the rest of the class on the pretest. In short, changes in I.Q. and SE performance were obtained by two different groups within the first KEEP kindergarten. Those who improved in I.Q. were from predominantly lower income families, while those who improved on the SERT showed a slight tendency to be from middle class families, as reflected by mother's employment and education. Those who improved on the SERT were, furthermore, already at or above average I.Q. on both the pre- and posttest administration of the WPPSI.

These data suggest that the significant and dramatic increases in I.Q. obtained by approximately half the class cannot be attributed to increased SE performance levels. The language factor did not hinder significant I.Q. gains by those initially at the lower end of the I.Q. distribution.

The SERT is scored for several classes of responses. The basic score is the number of accurate repetitions of 29 critical grammatical features. In addition, the child's response is scored for other forms of SE that are substituted and are appropriate linguistically. Moreover, the total SE correct and the total SE substitutions can be combined to form a measure of Total Appropriate SE.

In some cases, children respond to the SERT with Hawaii Creole English (HCE) substitutions. These HCE substitutions also vary in terms of appropriacy in the sense that a child may repeat a sentence using an HCE form that is

Table 5

Significant Correlates of Rep Test and HCE

	Sept. SERT (Pretest)	May SERT (Posttest)	SERT Change
Full Scale I.Q. Sept(Pretest)	.51	.64	.45
Verbal I.Q. Sept.	.53	.63	.40
Performance I.Q. Sept.	.44	.58	.45
Full Scale I.Q. May(Posttest)	.56 (27)	.70 (27)	.48 (27)
Verbal I.Q. May	.63 (27)	.76 (27)	.48 (27)
Performance I.Q. May		.43 (27)	
Full Scale Change SERT Test Change		.44	
Total On-Task		.42	
Total pages complete		.39	
1 month learning centers			
SE Rep Sept.		.91	
SE Rep May	.91		.44
Sample source	-.45	-.49	
Mother employed		.41	.44
Parents working	.43	.53	
Parents occup	.46	.53	
Father employed	.42 (25)	.44 (25)	
Father's occup	.49 (23)	.41 (23)	
Mother's occup	.43 (28)	.54 (28)	
Father's educ	.57 (23)	.54 (23)	
Mother's educ		.38 (28)	.51

equivalent to an SE form. These are scored as HCE appropriate substitutions.

Finally, by combining Total Appropriate SE and HCE, we obtain a measure of Total Language Performance. For heuristic purposes, we can assume that this Total Language score reflects, in some degree, the child's language performance level, in both dialects--SE and HCE.

The analysis of the correlations between the SERT change scores and these various categories of response to the SERT provides some further information on the structure of improvement on the SERT. Those children who improved on the SERT were more likely to have made HCE substitutions on the

pretest (r = .35, df=26). On the posttest, however, there is no relationship between HCE substitutions and SERT change scores. In other words, children who improved on the SERT switched from HCE substitutions on the pretest to accurate SE repetitions on the posttest. Therefore, it is uncertain whether their improvement is a function of increased familiarity with SE forms, or simply learning to use SE rather than HCE in school settings. The "improvement" on the SERT is probably a case of dialect switching cued by social factors, rather than a reflection of language learning.

There are only two correlations between SERT changes and socioeconomic factors. These two correlations are mother's educations (r=.51) and parent's employment status (r=.44). Although posttest scores correlated with all nine SES variables collected, and pretest scores correlated with six of the nine, change scores are less consistently related. Compared to those who show I.Q. change, the children who improve on the SERT are a somewhat more heterogeneous group with respect to SES factors; however, improvement does not alter the rank order of individuals in the groups since the pretest and posttest correlate very highly (r=.91, df=26). Quite independent of the change score, children who score high on the pre- and posttest are the same individuals, they have higher I.Q.s and are from middle-class families.

Apparently, children who improved on the SERT came from backgrounds in which HCE is an acceptable option. However, over the school year the children either learn that SE is the dialect of the teachers or they become more proficient in SE performance. In any event, these data we have reported suggest that dialect usage of entering kindergarteners does not predict school readiness as measured by I.Q. tests. It should be noted that there is no relationship between SERT change scores and the MRT, which is presumably a more direct measure of school readiness than I.Q.

Other Variables of Interest

A number of other variables analyzed proved of interest; for example, sex. Boys and girls were compared on 100 variables. There were 13 significant differences, by t test. Girls have higher I.Q. and achievement test scores in September, higher I.Q. scores in May, are on-task more as measured by daily observations, completed more pages in learning centers during a one-month sample period, have fathers with more education, and come from homes in which the father has less caretaking responsibilities.

Of the 100 variables, there were only three differences between children who attended preschool (n=13) and those that did not (n=12). They are I.Q. changes on Full, Verbal and Performance Scales. The fact that these critical variables were related to preschool attendance suggests more attention should be given to this finding than otherwise would be warranted by three significant differences out of 100 comparisons.

Summary

The pre- and posttest data suggest that KEEP did more for children whose incoming I.Q. and achievement scores were low and who were from families of lower socioeconomic status. Several explanations are plausible. First, the program is excellent and works well for children who need it most. Second, the kinds of pre- and posttest measures used are highly sensitive to changes in child/adult rapport, and thus, the increases are largely artifacts; that is, those who scored initially low were in fact capable of higher performance but because of social factors obtained depressed scores. Third, analysis of change scores is a difficult problem confounded by regression effects.

The data do indicate that the KEEP goals were met in an absolute sense. The children showed substantial improvement over the program year on meaningful

performance measures. Whether the results suggest that the KEEP approach is relatively superior awaits the collection of appropriate comparative data.

Implications for Further Research

1) Control groups must be obtained;
2) Findings must be verified by future KEEP classes;
3) The complex relationships among improvement, I.Q., and language ability must be studied further.